COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

An Analysis of the Supervised Learning Approach for Online Fraud Detection

Dhoma Harshavardhan Reddy [1], Dr.N Sirisha[2]

[1] B.Tech, Department of Computer Science and Engineering, MLR Institute of Technology, Dundigal, Hyderabad, India.
[2] M.Tech, Ph.D., Department of Computer Science and Engineering, MLR Institute of Technology, Dundigal, Hyderabad, India.

*Corresponding Author: [1] dhreddy2001@gmail.com, [2] nallashirisha@mlrinstitutions.ac.in

Abstract **- Illegal online financial transactions are now more sophisticated and global in scope, which costs both parties—customers and businesses. For fraud prevention and detection in the online setting, many different strategies have been proposed. While all of these techniques aim to detect and stop fraudulent online transactions, they differ in terms of their features, advantages, and disadvantages. This study assesses the current fraud detection research in this area to detect the employed algorithms and assessing in accordance with predetermined standards. The systematic quantitative literature review methodology was used to assess the research studies in the subject of online fraud detection. A hierarchical typology is created based on the supervised learning methods in scientific articles and their properties. Therefore, by integrating three selection criteria—accuracy, coverage, and costs—our research presents the best methods for identifying fraud in a novel approach.**

***Index Terms— Detection, Online fraud, Online transaction, Supervised Learning Algorithm*.**

1. Introduction

The threat of online exploitation is growing in the modern world as new technologies emerge. Fraud is defined as any illegal behaviour by a party that results in financial loss for a person or a business [1]. Fraud can happen in many different contexts and ways. Understanding and using fraud analytics is an effective strategy to combat fraud and the criminals [2]. It aids in the prevention, detection, and mitigation of fraud. The topic of identifying fraud in the area of online transactions is covered in this essay. Fraud detection is a process through which any harmful conduct that has already occurred and caused a loss of any type to the target entity can be proactively discovered and detected [1]. Every day, millions of transactions are made online, and each of these transactions is vulnerable to different types of fraud. These transactions include any online financial exchanges. There are many different types of fraudulent online transactions, including fraud involving credit card transactions, bank statements, insurance, and automated banking activities [3]. With a publicly accessible credit card dataset, the purpose of this study is to elaborate on fraud detection using supervised machine learning algorithms.

Due to the vastness of the databases, manual dependency is impossible to evaluate and analyse. Consequently, the idea of machine learning was presented. In the study, the input data were subjected to the supervised learning technique [4]. In this method, the system is shown training data as well as the intended result. Following calculations, it intelligently determines the best answer. The effectiveness of each strategy may be calculated and compared with one another because the target output was already known when the results were to be calculated. The dataset in the study that has been given has been processed using supervised learning strategies. Depending on whether the fraudulent transactions that are produced match those in the dataset, the findings acquired have demonstrated the accuracy of each technique. These methods include random forests, support vector machines, decision trees, naive Bayes, nearest neighbors, and logistic regression methods. After being applied to the dataset, these machine learning methods produced expected outcomes with an accuracy rate of >90%.

The dependence on online payments and e-commerce has increased over the past few decades. The majority of businesses and individuals are experiencing significant financial losses as a result of the worldwide growth in illicit attempts to conduct online transactions as the field of information technology continues to advance and get better over time [5]. "The misuse of a profit organization system without necessarily resulting to immediate legal consequences" is how fraud is characterized. On the basis of algorithms and analytical tools, complex decision-making systems have been created. These have the capacity to learn from past mistakes and develop patterns that enable proactive detection of possibly fraudulent transactions. This study seeks to present a summary of current strategies for fraudulent detection based on the most notable standards after reviewing

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

several significant research publications from the past few years. The program should process a significant amount of transaction data with great accuracy and high precision. The algorithm should facilitate obtaining high coverage for fraud with low false positive rates.

2. Related Work

Kamaruddin and Ravi [6] created a hybrid method of optimization by auto-associative neural network and particle swarms on the Spark computing platform for one-class classification. Gómez et al. [7] used ANN to detect transactions with fraudulent credit cards and reduce data imbalance. They also calculated the expenses related to the results. Santiago et al. [8] SVM classifier was used to determine whether or not a credit card dataset transaction was fraudulent. Self-organizing maps' grouping and filtering capabilities have been used by Quah and Sriganesh [9] to develop a method for identifying credit card fraud. Bhattacharyya et al. [10] investigate the methods of integrating SVM, Random Forest (RF), and linear regression allowed them to detect fraud credit cards.

Panigrahi et al. [11] successfully detected credit card fraud by using four techniques: Dempster-Shafer adder, rule-based filter, Bayesian learning and database of transaction histories. They initially identified the questionable transactions by looking at how they deviated from the expected trend. After that, the sum of these cases was computed to produce a first impression. After categorizing the transactions based on the level of fraud, Bayesian learning is used to determine whether or not the belief is valid based on past transaction data. Halvaiee and Akbari [12] created the fraud detection with an artificial immune system algorithm to identify credit card fraud. In an unbalanced dataset, implemented supervised vector classifier and logistic regression perform better at detecting credit card fraud [13], in big data analytics [14]. Using Machine Learning Online Fraud Detection studied in [15], [16]. Despite the fact that many studies have been done on the subject, few have compared and used several supervised learning algorithms to detect fraud in online transactions.

3. Online Fraud Detection

For all bank institutions, managing financial crime has become a severe problem due to online banking fraud. Due to the growth and development of sophisticated and inventive online fraudulent, including malware infection, phishing schemes, and ghost websites, it is becoming increasingly difficult and results in significant losses. The majority of clients rarely routinely review their online banking history, making it difficult for them to promptly identify and report fraudulent transactions. As a result, there is extremely little chance of recovering losses. The protection of e-commerce, credit card transactions, retail, insurance, communications, computer infiltration, etc. are all significantly hampered by these qualities. When used specifically for detecting fraud in online banking, these existing approaches fare badly in terms of accuracy and/or efficiency. For instance, detecting credit card fraud frequently relies on identifying certain customer or group behaviour patterns, whereas internet banking transactions involving fraud are extremely dynamic and closely resemble real customer behaviour. Although they need a lot of training data and entire attack logs as proof, some intrusion detection techniques work effectively in dynamic computing environments. However, it might be challenging to determine whether an online banking transaction is fraudulent without clear evidence.

Online fraud refers to theft and fraud that is carried out through any type of payment method, including cards, online transactions, etc [17]. The following are various types of online fraudulent.

- Clean: It is a form of fraud where the perpetrator can utilize the merchant's checks by posing as an authorized user.
- Account Takeover: A fraudster links the account of a legal user to his credit card.
- Friendly: The merchant has received a chargeback after a legitimate user disputes the transaction. The term "Chargeback fraud" is often used.
- Identity: Identity fraud refers to the acquisition of sensitive personal information such as a passport or account number through impersonation.
- Affiliate: Affiliate fraud is the online use of company information for personal gain.
- Re-shipping: The practice of using a mule or recruited individual to re-ship goods that have been bought with fraudulent credit cards.
- Botnets: A robot or machine that conducts online transactions using a fake credit card's physical location. The transaction appears to be legitimate as a result of the shared geographic location.
- Phishing: Fraudsters send emails that appear to be from legitimate sources in order to collect sensitive data from verified users.
- Whaling: Similar to phishing, but the target is a predetermined group of customers who are members of a successful online business.
- Phishing: Customers are directed to illegitimate websites when purchasing online, when sensitive data may be collected.

- Triangulation: Through third-party online auction or ticketing sites, credit card information is acquired from verified

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

customers.

The misuse of Internet banking resources and web technology in the digital world, and the abuse of both of these in the physical world—reflects the misuse of interactions between those resources [5]. Hence abusing the exchange of goods and services in the real world.

In the same research, we discover that the following traits and difficulties are common to online fraud detection:

- The data set is enormous and severely unbalanced; for instance, just 5 examples of fraud were included in a very large data set of more than 300000 transactions in a single day, making it hard to identify the few instances of fraud among the vast number of valid transactions.
- The fraud behaviour is dynamic; as information technology develops on a daily basis, fraudsters constantly improve the methods they use to get around online banking security.
- There are many different types of customer behaviour patterns; in this situation, fraudsters frequently imitate real customer behaviour.
- Additionally, they frequently alter their behaviour to keep up with developments in detection of fraud. All of these make it challenging to define fraud and even more challenging to separate it from sincere action.

Therefore the consumer accesses the same banking system every time they use the online banking service, which makes it possible to characterize typical sequences and spot red flags in banking fraud on online transactions.

The aforementioned aspects make fraud detection particularly difficult, which is why numerous machine-learning algorithms have been created to address this issue [18]. The algorithm was motivated by the need to examine a data collection of around 15 million actual online banking transactions from 2011 to 2013 in order to distinguish fraudulent from legitimate transactions. The algorithm has somewhat low false positive rates and achieved high true positive rates which supported the findings that it is particularly good at spotting anomalies.

4. Research Methodology

The objective of the research is to evaluate and categorize supervised -learning methods that can effectively identify bank fraud in the online setting while also meeting the following requirements: minimal costs, high accuracy and high coverage.

As a result, a meta-analysis was carried out on a range of particular articles (conference papers and peer reviewed journal articles) from 2010 to the present that satisfied the set criteria (low costs, high accuracy, high coverage) and followed these descriptors are from articles' titles and abstracts: identifying bank fraud, identifying bank fraud online, identifying bank fraud using supervised learning. The descriptors that best define the most popular supervised learning approaches used to identify fraud in online transactions were chosen depending on the subject of this article. The articles were obtained from sources such as Springer Link, Science Direct, IEEE Xplorer Digital Library, ACM Digital Library, etc. after the descriptions.

A. Supervised Learning Approach

ML algorithms treat each instance of a dataset as a grouping of features. Depending on their nature, these traits could be continuous, categorical, or binary. If the examples are labeled, this approach of learning is referred to as supervised learning [19]. After being trained using supervised learning on labeled data, the model is evaluated on unlabeled data. Datasets are first gathered, then split into test and training sets, and preprocessed as one of the fundamental design.

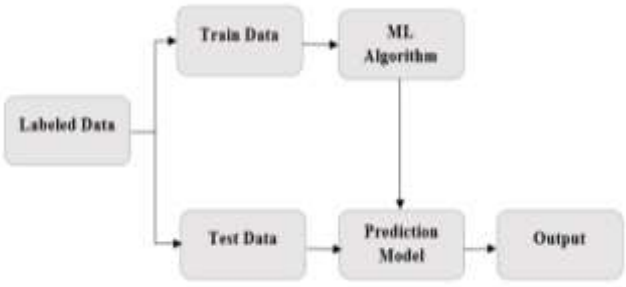

Fig. 1. Basic Architecture of Supervised Learning

The framework is trained to recognize the attributes related to each label after being fed the extracted features into an algorithm. The model is then given the test data, and as shown in Fig. 1, it uses the predicted labels to produce predictions on the test data. Classification and regression are the two main subtypes of supervised learning.

Finding and evaluating works on the issue under consideration is the main objective of this literature review. By offering an overview of Supervised Learning Approaches and Algorithms, the metrics employed to assess each supervised learning model's performance, and a comparative study of the accuracy of each supervised learning model, this review will help researchers discover future research fields. As everyone is aware, a significant portion of internet usage comprises browsing (visiting websites). This is one of the main motives used by phishers to target users and steal their personal information.

COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

Various types of algorithms are used to stop various types of attacks by Supervised learning algorithms which is depicted in table 1. The study's comparative methodology's system architecture is shown in Fig. 2.

Table I. Description of the Supervised Learning Methods

Algorithm	Key features	Implementation in fraud analytics	Advantages	Disadvantages
Naïve Bayes	This methodology uses the Bayes theorem and is a classification method. The dataset's data points are all regarded as independent.	Naive Bayes predicts if a particular transaction is fraudulent or not.	This approach requires less training data and will converge rapidly.	A reasonably strong assumption is made by the Naive Bayes classifier.
Logistic regression	In logistic regression, it is possible to predict the probability that a particular instance would fall into either class "1" or class "0." The sigmoid function model is then used to convert the output so that it can only take discrete values of 0 or 1.	This study makes a prediction as to whether a specific instance when it does not correlate to fraud (class "0") and if data belongs to fraud (class "1").	Discrete variables are predicted using this technique.	The linear regression is not used to forecast binary variables since it produces results in a continuous range.
Random forest	An approach to ensemble learning is random forest. It randomly chooses a portion of the training set during training and builds a decision tree for that subset. In the training phase, this procedure is performed numerous times, producing various decision trees.	This algorithm constructs numerous decision trees using a subset of the dataset that is randomly chosen, and the outcome determines specific fraudulent transaction as ("0") and for not as ("1").	The most accurate learning system, random forest, creates very accurate classifiers.	If the dependent and independent variables have a linear relationship, then this strategy is inaccurate.
Nearest neighbors	Using the labels given to the k-nearest training illustrations, this approach categorizes a dataset instance.	Each instance to test using the Euclidian distance metric. The model will come to the conclusion that the given transaction is fraudulent is “1” and not is “0”.	It manages multi-class cases.	Useful distance function is required as it has a significant computation cost.
Decision trees	Each internal node of the tree in a decision tree reflects a choice and divides the tree into branches based on a condition. A judgment regarding the label of the test instance is represented by the tree's leaf node.	The test instance is marked by the leaf node as either fraudulent ("1") or not ("0") During the training process, the model optimizes the tree's structure. Recursive binary splitting is can be employed to train an algorithm with minimal cost.	The main benefit of this approach is that the performance of the tree is unaffected by nonlinear interactions between parameters.	Any modification to a dataset's data value causes the decision tree to become unstable.
SVM	Support vector	28 features make up the	It is beneficial when a	The greatest

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

	machines (SVM) create a hyperplane—a surface of dimension n-1 in a space of n dimensions—during the training process that best classifies the training data. The labels for the test data are then determined using the hyperplane. SVM kernels transform the data in a nonlinear way. The feature space in which SVM functions can be changed by kernels.	dataset. A 27-dimensional hyperplane will be built via SVM. The data will be optimally split into two parts by this hyperplane, with one portion containing predominantly fraud instances and the other containing non-fraud examples.	hyperplane in the original feature space cannot separate two data points. If so, kernels can change the feature space into a higher-dimensional space in which a hyperplane divides the data points. The feature space is transformed by the RBF kernel into an infinite-dimensional space.	drawback of SVM is the choice of kernel function parameter [20].

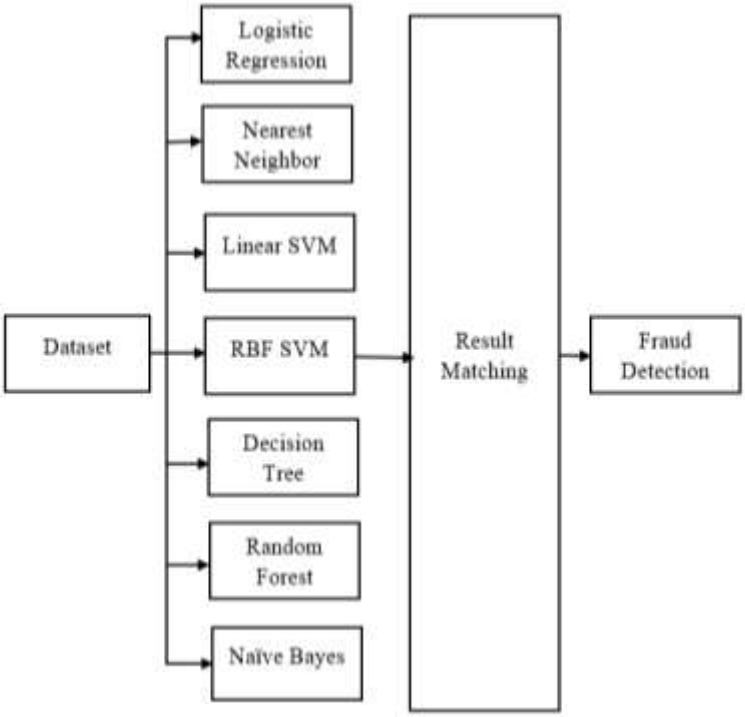

Fig. 2. The study's comparative methodology's system architecture

According to the study based on fraud types, 56 articles contain credit frauds, financial frauds, and e-commerce frauds. There are many supervised learning approaches stated in these articles that have been analysed with a range of numbers, from 1 to a maximum of 6. The Logistic regression, KNN, SVM, Random forest, Bayesian and Decision Tree are most important techniques that are used in more than 20 articles. Some eliminated due to their low frequency from the results section. Hence out of these 48 scientific articles utilised data sets to support the study. We primarily focused on papers that utilized a significant amount of data and had positive outcomes for the three primary criteria: high coverage, high accuracy, and low prices. DT, SVM, KNN and Naive Bayes all achieved medium and high coverage rates on datasets with more than 1 million data records or online transactions as well as those with thousands to several hundred thousand records [21-33], [34] and [35]. On more than 1 million data records or online transactions as well as on those that provided thousands to several hundred thousand data records or online transactions, DT, SVM, KNN and Naive Bayes all shown high and medium accuracy rates [21], [24], [26-29], [31] and [32]. Costs were considerable for all methods used on the massive amounts of data records.

The main purposes of supervised learning algorithms, which include the SVM, Decision Tree, Bayesian network, Logistic regression, Nave Bayes, K-NN and Random forest are accurate classification and prediction. The most popular strategies now are supervised learning algorithms, which have the disadvantage of being more expensive than fraud but also offer high accuracy and high coverage. Figure 3 shows the publication of Supervised Learning Approach from 2010-2022.

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

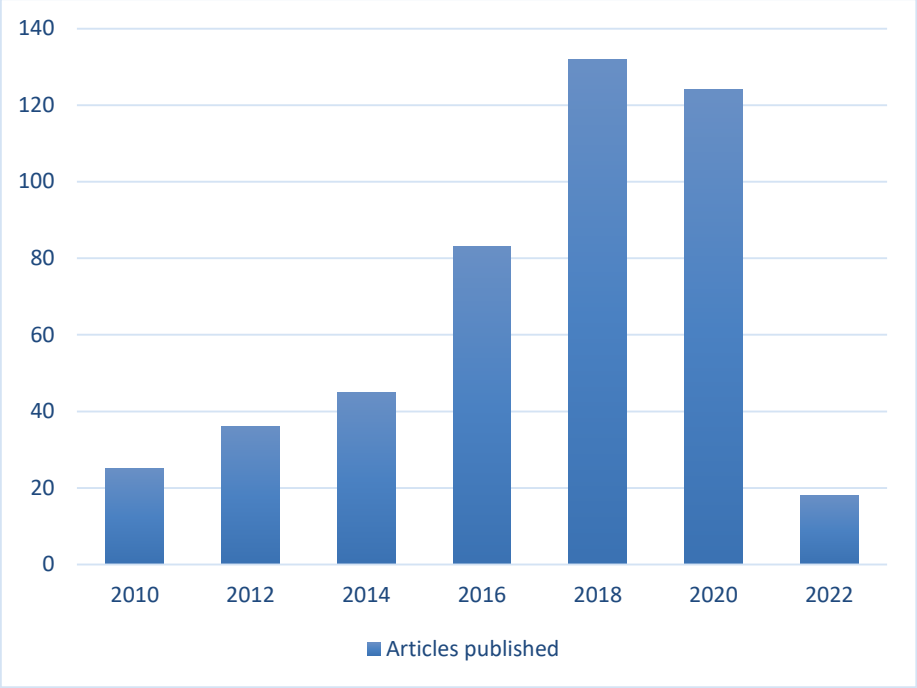

Fig. 3. Publication of Supervised Learning Approach from 2010-2022

5. Research Results

In this study, we conducted a comparative analysis of the most frequently used supervised learning algorithms for detecting online fraud frequency, and on certain criteria: High accuracy should be achieved by the algorithm when processing massive amounts of data transaction. The algorithm must facilitate obtaining high coverage for fraud with low false positive rates. In terms of time and cost efficiency, the algorithm should benefit both corporations and individual users.

The algorithms in Table 3 were categorized according to the positive and negative cases that a classifier accurately predicted, which metrics may be found in the following formulas:

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

$$FPR = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

$$TNR = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$TPR = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}$$

True Negatives (TN) and True positives (TP) are the negative and positive occurrences that a classifier accurately predicts. False Negatives (FN) and False positives (FP) are the terms for the events that were mistakenly categorised. A classification of the techniques is aided by the notions of TN, TP, FN and FP rates, Accuracy and Precision, that will be based on these rates. Since it's crucial to have algorithms, the correctness in the application of these metrics in the three criteria. Since we lacked access to the datasets needed we utilize these dataset formulas in future work to see if the machine-learning approaches we obtain have the accuracy claimed in the reviewed literature.

In the following section, we compare the outcomes of the most widely used online fraud detection techniques in the criteria– costs, coverage, and accuracy, where 1 denotes low, 2 denotes medium, and 3 denotes high. This is done in accordance with the analysis of the reviewed articles and the aforementioned metrics.

COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING

Computational Intelligence and Machine Learning
Vol-3 Issue-2, October 2022, PP.47-56
ISSN : 2582-7464 (Online)

Table II. Algorithm classification using the specified criteria and frequency

Supervised Learning algorithm	Costs	Coverage	Frequency	Accuracy
Decision Tree (DT)	3	2	38%	2
Support Vector Machine (SVM)	3	3	34%	3
K-nearest Neighbor (KNN)	3	2	20%	2
Logistic regression	2	2	16%	3
Random forest	2	2	16%	3
Naïve Bayes	3	2	14%	2
Bayesian Network	3	2	16%	3

As results demonstrate, the Bayesian Network, SVM, fraud detection systems have very high accuracy with 100% true positives but have the drawback of significant processing costs when processing with large datasets. In a different perspective, while processing massive datasets, genetic algorithms and AIS offer medium accuracy with minimal costs. DT, KNN, and Naive Bayes, for example, present medium coverage with the drawback of high costs, and medium accuracy when compared to other algorithms.

According to a review of the literature [30] and [33, 34], supervised learning approaches seem to be the most popular methods for identifying online fraud, despite their high cost. Additionally, according to the literature [24, 35], credit card transactions are the most common fraudulent ones in the online environment, which is supported by our data in table 2. As a result, the classification based on the chosen criteria attempts to create more reliable and effective fraud detection systems, which should also take into account variables like bank and customer behaviour, risk levels, geographical locations, and so forth. In conclusion, the results indicate that supervised approaches have good accuracy and coverage with the drawback of high costs. It is crucial to remember that these results are positive because the datasets utilized were quite uneven and had a large number of negative occurrences. Lack of access to the data sets utilized in the examined studies to identify the characteristics of the approaches was one of our research's limitations. Finally, we believe that our research combines three selection criteria—accuracy, coverage, and costs—to highlight the most effective fraud detection methods in a novel way. The comparison of Supervised Learning Methods in fraud detections is depicted in table 3.

Table III. Comparison of Supervised Learning Methods in fraud detections

Ref	**Methods**	**Dataset used**	**Performance Metrics**	**Accuracy**
[36]	Logistic-regression, Deep learning	Nigeria bank	It enhanced the quality of real-world transaction entries.	95% for detecting fraudulent transactions
[37]	SVM, Decision Trees	Nation banks credit card warehouse	Accuracy	SVM is outperformed by decision trees.
[38]	NB, KNN, D.T., L.R., CFLANN	Europeans cardholders	CFLANN reduces mean-squared error	detecting fraudulent transactions 97.56%
[39]	Neural networks and SVM	Chinese financial institution	The CCFD's overall performance is improved	99.21% accuracy and recall of 95.20%
[40]	RTRF, CRF, RF-1, RF-2,	Chinese E-commerce firm	In comparison to other D.T. algorithms, R.F. performs well.	96.77% accuracy and 89.46% precision
[41]	logistic regression, Random forest (R.F.), D.T., SVM, KNN	Datasets from Europe credit card	Classification techniques result in increased prediction accuracy	Compared to random forest, a better level of overall accuracy is achieved.
[42]	KNN, SVM	Datasets from financial institutions	It assists in categorizing transactions in actual situations.	72% accuracy by KNN 91% by SVM
[43]	SVM	Banks datasets	SVM avoids overfitting	SVM outperforms the hybrid B.P. model in terms of performance.

[44]	KNN, SVM, Naive Bayes (NB)	UCSD FICO datasets	The less significant change has little effect on how the model is implemented.	10% accuracy by KNN, 15% by NB and 20% by SVM
[45]	KNN	UCI websites	Predictive models are not required for prior classification.	72% accuracy by KNN
[46]	SVM, Logistic Regression and Random Forests	From a study on ANN optimized by Genetic Algorithms (GAs) to detect fraud from international credit card operations	Accuracy, specificity, sensitivity, F-measure, precision, G-mean, wtdAcc.	Better performance overall than Random Forest. Higher performance with different datasets and logistic regression
[47]	K-NN	Authentic data from a private bank	F-measure, Recall, Accuracy, Specificity, Precision	Performance is assessed based on the metrics.

6. Conclusion

Our study indicates that the literature has given the most attention regarding the issue of online credit card fraud, despite the fact that there are a number of important issues that the researchers have not focused on in great detail, such as pagejacking, online intellectual property theft, wire-transfer fraud and false money orders. The categorization of the algorithms revealed that the supervised learning techniques—SVM and decision tree—produced the greatest results in terms of coverage and accuracy. These three algorithms also received the most calls in the articles we evaluated, showing that they deliver the greatest outcomes. Therefore, in terms of research direction, it is preferable to explore into potential algorithmic enhancements that could expand to other forms of online fraud transactions with high coverage, high accuracy, and low costs.

REFERENCES

[1] A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” Journal of Network and Computer Applications, vol. 68, pp. 90-113, 2016.

[2] S. Makki, R. Haque, Y. Taher, Z. Assaghir, G. Ditzler, M. S. Hacid and H. Zeineddine, “Fraud analysis approaches in the age of big data-a review of state of the art”. In 2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS * W), pp. 243-250, 2017.

[3] West, Jarrod, and Bhattacharya, Maumita. “Intelligent financial fraud detection: a comprehensive review. Computers & Security, vol. 57, pp. 47-66, 2016.

[4] B. Baesens, V. Van Vlasselaer and W. Verbeke, “Fraud analytics using descriptive, predictive, and social network techniques: A guide to data science for fraud detection,” Wiley, 2015.

[5] W. Wei, J. Li, L. Cao, Y. Ou and J. Chen, “Effective detection of sophisticated online banking fraud on extremely imbalanced data”, World Wide Web, 2013.

[6] S. Kamaruddin and V. Ravi, “Credit card fraud detection using big data analytics: use of psoaann based one-class classification. In Proceedings of the International Conference on Informatics and Analytics, p. 33. ACM, 2016.

[7] J. A. Gómez, J. Arévalo, R. Paredes and J. Nin, “End-to-end neural network architecture for fraud scoring in card payments,” Pattern Recognition Letters, vol. 105, pp. 175-181, 2018.

[8] G. P. Santiago, A. Pereira and R. Hirata Jr, “A modeling approach for credit card fraud detection in electronic payment services”, In Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp. 2328-2331, ACM, 2015.

[9] J. T. Quah and M. Sriganesh, “Real-time credit card fraud detection using computational intelligence”, Expert Systems with Applications, vol. 35(4), pp. 1721-1732, 2008.

[10] S. Bhattacharyya, S. Jha, K. Tharakunnel and J. C. Westland, “Data mining for credit card fraud: A comparative study,” Decision Support Systems, vol. 50(3), pp. 602-613, 2011.

[11] S. Panigrahi, A. Kundu, S. Sural and A. K. Majumdar, “Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning”, Information Fusion, vol. 10(4), pp. 354-363, 2009.

[12] N. S. Halvaiee and M. K. Akbari, “A novel model for credit card fraud detection using artificial immune systems”, Applied Soft Computing, vol. 24, pp. 40-49, 2014.

[13] P. Verma and P. Tyagi, "Analysis of Supervised Machine Learning Algorithms in the Context of Fraud Detection. ECS Transactions, vol. 107(1), pp. 7189, 2022.

[14] A. Pandey, H. Jaiswal, A. Vij and T. Mehrotra, "Case Study on Online Fraud Detection using Machine Learning," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 48-52, doi: 10.1109/ICACITE53722.2022.9823538.

[15] Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. Human-Centric Intelligent Systems, 1-14.

[16] Y. K. Saheed, U. A. Baba and M. A. Raji, "Big Data Analytics for Credit Card Fraud Detection Using Supervised Machine Learning Models", In Big Data Analytics in the Insurance Market (pp. 31-56). Emerald Publishing Limited, 2022.

[17] T. P. Rani, Magilan Saravanan, Ashish Kumar Sahu, K. Martin Sagayam, and Ahmed A. Elngar. "Predicting Online Fraudulent Transactions Using Machine Learning," 2022.

[18] S. J. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach", 3rd edition. Prentice Hall, 2010.

[19] I. Muhammad and Z. Yan, "Supervised Machine Learning Approaches A Survey," ICTACT Journal on Soft Computing, vol.5(3), 2015.

[20] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," 2008.

[21] K. Navanshu and S. Y. Sait, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models", International Journal of Pure and Applied Mathematics, Volume 118, No. 20, pp. 825-838, 2018

[22] K. R. Seeja and Z. Masoumeh, "Fraud miner. A novel credit card fraud detection model based on frequent item set mining", The Scientific World Journal, Article ID 252797, 2014

[23] D. Panaro, E. Riccomagno and F. Malfanti, "A Fraud Detection Algorithm For Online Banking", 2015

[24] J. K.-F. Pun, "Improving Credit Card Fraud Detection using a Meta-Learning strategy", a thesis submitted in conformity with the requirements for the degree of Master of Applied Science Graduate Department of Chemical Engineering and Applied Chemistry University of Toronto, 2011

[25] Y. Sahin, S. Bulkan and E. Duman, "A cost-sensitive Decision Tree Approach for Fraud Detection", Expert Systemts with Applications, Volume 40, pp 5916 – 5923, 2013

[26] Q. Lu and C. Ju, "Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine", Journal of Convergence Information Technology, pp. 62, 2011

[27] D. D. Patil, V. M. Wadhai and J. A. Gokhale, "Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy", International Journal of Computer Applications, Volume 11, No. 2, pp. 23 – 30, 2010

[28] S. Soltaniziba, M. A. Balafar, "The Study of Fraud Detection in Financial and Credit Institutions with Real Data", Computer Science and Engineering, Volume 5, No 3, pp 30-36, 2015

[29] M. Carminati, R. Caron, F. Maggi, I. Epifani and S. Zanero, "Banksealer: A decision support system for online banking fraud analysis and investigation", Computers and Security, 53:175–186, 2015

[30] A. Sinha and S. Mokha, "Classification and Fraud Detection in Finance Industry", International Journal of Computer Applications, Volume 176, No 3, 2017

[31] S. Kiran, N. Kumar, J. Guru, D. Katariya, R. Kumar and M. Sharma, "Credit card fraud detection using Naïve Bayes model based and KNN classifier", International Journal of Advance Research, Ideas and Innovations in Technoloy, Volume 4, Issue 3, 2018

[32] A. C. Bahnsen, S. Villegas, D. Aouada and B. Ottersten, "Fraud Detection by Stacking Cost-Sensitive Decision Trees", Data Science for Cyber-Security, 2017

[33] R. Banerjee, G. Bourla, S. Chen, M. Kashyap, S. Purohit and J. Battipaglia, "Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection", New Jersey's Governor's School of Engineering and Technology, 2018

[34] J. R. Gaikwad, A. B. Deshmane, H. V. Somavanshi, S. V. Patil and R. A. Badgujar, "Credit Card Fraud Detection using Decision Tree Induction Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 4, No. 6, 2014.

[35] C. Mishra, D. Lal Gupta and R. Singh, "Credit Card Fraud Identification Using Artificial Neural Networks", International Journal of Computer Systems, Volume 04, Issue 07, 2017.

[36] F. N. Ogwueleka, "Data mining application in credit card fraud detection system," J Eng Sci Technol. Vol. 6, pp. 311–22, 2011.

[37] S. Yusuf, E. Duman, "Detecting credit card fraud by decision trees and support vector machines," IMECS 2011- International multiconference of Engineers and Computer Scientists 2011, 1, 442-447, 2011.

[38] T. Jemima Jebaseeli, R. Venkatesan, K. Ramalakshami, "Fraud detection for credit card transactions using random forest algorithm," Singapore: Springer; 2020.

[39] R. Patidar and L. Sharma, "Credit card fraud detection using neural network," Int. J. Soft Comput Eng (USCE), vol. 1, pp. 32-38, 2011.

[40] O. Vynokurova, D. Peleshko, O. Bondarenko, V. Ilyasow, V. Serzhantow, and M. Peleshko, "Hybrid machine learning system for solving fraud detection tasks. In: 2020 IEEE third international conference on data stream mining & processing (DSMP), IEEE; pp. 1 – 5, 2020.

[41] S. Dhankhad, E. Mohammed and B. Far, "Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study," In: IEEE, pp. 122-125, 2018.

[42] H. Zhu, G. Liu, M. Zhou, Y. Xie, A. Abusorrah and Q. Kang, "Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection," Neurocomputing. Vol. 407, pp. 50-62, 2020. https://doi.org/10.101i/j-neucom.2020.04.078.

[43] M. K. Mishra and R. Dash, "A comparative study of Chebyshev functional link artifcial neural network, multi-layer perceptron and decision tree for credit card fraud detection," In: IEEE, p. 228–233, 2014.

[44] V. Sobanadevi and G. Ravi, "Handling data imbalance using a beterogeneous bagging-based stacked ensemble (HBSE) for credit card fraud detection", Singapore: Springer; 2020.

[45] F. Itoo and S. S. Meenakshi, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," Int J Inf Technol. 2020;13:1503–11. https://doi.org/10.1007/s41870-020-00430-y.

[46] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, 50, 602-613, 2011.

[47] M. Sanaz, S. Mehdi, "Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors.," 2018.